DOES VENTURE CAPITAL ACCELERATE THE INVENTIVE ACTIVITIES OF FRONTIER TECHNOLOGIES? NEW EVIDENCE FROM ARTIFICIAL INTELLIGENCE

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Chenyang Cao

August 2019

ABSTRACT

The primary purpose of this thesis is to update the established findings on the complementary relationship between venture capital (VC) investments and federal R&D funds in stimulating inventive activities concerning about artificial intelligence (AI) patenting. Using panel data from 1990 to 2014 on top 150 MSAs/CMSAs, OLS regression results with fixed-effect estimates implied that i) the supply of VC investment increased the rates of patents, AI patents, and IT patents; ii) calculations of elasticity indicated that federal R&D funds significantly influenced the development of frontier technological inventive activities like AI instead of the traditional information technologies; iii) the complementary effect between VC investments and federal R&D funds consistently exited on the growth of patents, AI patents, and IT patents; and iv) the elasticity of VC investments reached the largest on the rates of AI patents when there is a high presence of federal R&D funds within an MSA/CMSA. Besides, the OLS results remained robust after altering the number of top MSA/CMSA selected, relaxing the searching criteria of AI patents, and changing the patents selection with a subset that contains only USPC classification codes, respectively. However, IV estimates didn’t show robustness to the OLS results with fixed effects estimates.

BIOGRAPHICAL SKETCH

Chenyang is a Master of Science student in the field of Applied Economics and Management at Cornell University. Before started her journey at Cornell, Chenyang graduated as Bachelor of Science in Economics from Renmin University of China (RUC). Chenyang's thesis was supervised by Dr. Chris Forman, and Dr. Panle Jia Barwick served as her committee minor member. Her research interests lie in firm management practices and technological innovation.

This document is dedicated to all Cornell graduate students.

ACKNOWLEDGEMENTS

I am very much indebted to my committee chairperson, an incredible advisor, Dr. Chris Forman for his generous advice, patience, and insights. I also want to especially thank my minor committee member Dr. Panle Jia Barwick for her trust, guidance, and motivation. Although it is not directly related to this thesis, I am deeply grateful to the chance of working with Dr. Jennifer Ifft at Cornell University, Mr. Raian Divanbeigi and Dr. Marcio Cruz from the World Bank Group. Please accept my sincere gratitude to you all for enlightening me and being role models in all aspects of my career and personality development.

For my parents and my lover who have always been my strong and solid support. For my friends, mentors, and cohorts who have been bearing with me all the way long. For professors, librarians, Mr. Ken Redd at NACUBO, and many more people who have kindly helped a person from nowhere (for her annoying questions and data issues). Nothing would be accomplished without you all.

I also owe a big thank to Cornell University, where I have grown, explored, and discovered many of the most meaningful things in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

It is widely understood that there exhibits a positive correlation between innovation and productivity growth [1] [17]. The faith has been a strong reason for firms and governments to allocate a large portion of financial resources on inventive activities. Previous findings have shown the support of public research funding and the accessibility of private financial resources are often related to the progress on scientific inventions [16][21] . With the popularity of frontier technologies like artificial intelligence (AI), it is of growing interest in what contributes to the prosperity of AI invention. Specifically, there are many questions to be asked: Does the geographical distribution of investments match with the emerging areas of AI technologies? What kinds of public and private investments are indispensable to AI invention? Does private or public capital flows primary influence on AI invention? Will the coexisting of both parties makes inventive outcomes better off?

In this thesis, I intend to study the relationship between public and private research funding concerning AI-related patenting behavior. Extending this line of inquiry, I started by choosing proper measurements of AI invention activities and financial investments. Based on the empirical models and evidence that show a positive relationship between patents and R&D [8] [11] [12] [22], this thesis adopted the count of patents as the measurement of invention activities. However, the definition of AI patents has not achieved a certain consensus. Motivated by the work of Cockburn, Henderson, and Stern (2018) [6], this thesis applied their criteria of AI patents selection, which uses USPC code and keywords found in peer-reviewed and public-domain literature on AI. The thesis further developed the analysis in depth by building up a comparison group, non-AI IT patents, which contains the most related but non-AI patents. The construction of a comparison group was inspired by the paper of Forman, Goldfarb, and Greenstein (2016) [9]. Their paper extensively documented different regional growth patterns of ICT and non-ICT patents. Whereas, this thesis modified this comparison and construed a group of non-AI IT patents using the HJT class [10] that AI USPC sub-classes are indented under. For the measurement of AI-related investments, I referred the fact from Rin, Hellmann, and Puri's paper on venture capital research survey [19] that technology companies especially start-ups and enterprises have strong propensity to acquire financial support from Venture Capital (VC) investments. Meanwhile, prior work has assessed the role of VC in entrepreneurship [14] [20] [21]. Their findings consistently suggest that increasing in VC investment activities is associated with a significantly growth of patenting rates. Apart from the role of VC, public research funding is often recognized as another factor that influences technological development[21] [24]. Moreover, the comparison between public and private research has been one of the fundamental issues discussed in the field of policy design [2] [16] [20].

To connect all the dots into a line, this thesis will focus on exploring: i) if there exists a positive relationship between VC investments and growth in patents as suggested in previous findings [20][24] ; ii) if the relationship between VC investments and patenting rates are likely to be stronger for frontier technologies, such as AI; iii) if there exists a complementary relationship between public and private RD on patenting rates as suggested in previous findings [20]; iv) if there exists a relationship between public and private RD and its potential complementarity for AI and non-AI software patents.

As a quick reminder, this thesis will be arranged in the following order: i) a literature review discussing studies on regional pattern of patenting and relationships between public and private research funding; ii) data and methodology used to study the comparison and complementary effect between VC investments and public research funding on three selections of patenting; iii) a summary of the main results with discussions of problems and future work; iv) the appendix consisting of descriptive statistics and results from robustness checks.

CHAPTER 2

LITERATURE REVIEW

2.1 Comparison of patenting, AI patenting, and IT patenting

Previously, several investigations have been conducted on exploring the regional patterns of patenting and its dynamics in accelerating technology commercialization [7]. For instance, Powell (2002) [18] identifies the cluster of innovation as an essential factor of production in the commercial field of biotechnology. Carlino and Kerr (2015) [4] reviews the connections between the regional patterns of agglomeration and innovation using patents as a principle measurement; Forman, Goldfarb, and Greenstein (2016) [9] find the dominance of patenting in the Bay area. However, less literature is found to document the patenting pattern of a specific technology with details. This thesis applied the criteria of AI patents selection, which developed by Cockburn, Henderson, and Stern [6] using USPC code and keywords found in peer-reviewed and public-domain literature on AI. To update the findings on regional patterns, figures 2.1, 2.2, and 2.3 plotted the fractions of utility patents, non-AI IT patents, and AI patents selections in top 10 MSA/CMSA over the decade from 1990 to 2014 as a demonstrate of different regional patenting patterns applied over the period. Aligned with findings from Forman, Goldfarb, and Greenstein (2016) [9], the graphs in this thesis revealed a substantial agglomeration of invention, especially technology-related invention, in urban areas. It could also be observed from figure 2.3 that the Bay area, New York metro area, and Boston metro area are the top three clusters of AI inventions.

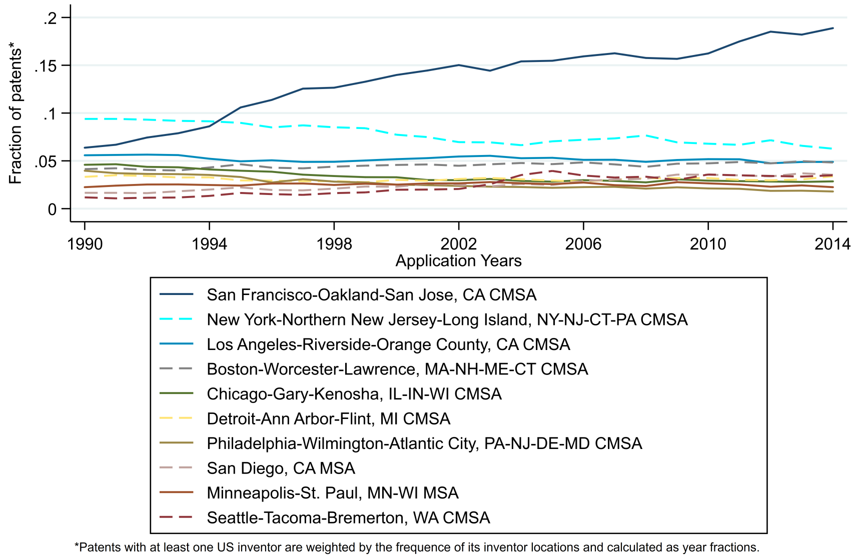

Figure 2.1: Fraction of Utility Patents in Top 10 MSAs/CMSAs

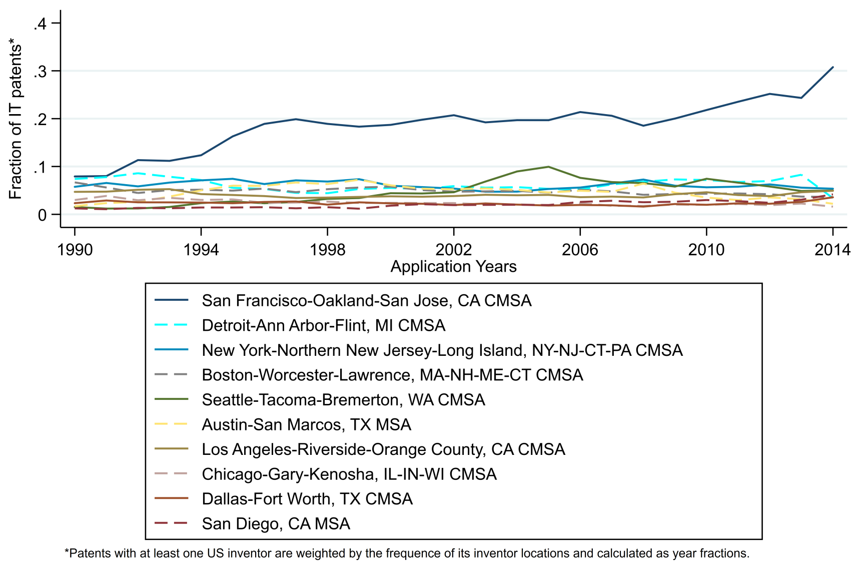

Figure 2.2: Fraction of non-AI IT Patents in Top 10 MSAs/CMSAs

Furthermore, both invention and investments tend to be localized. Starting from late 90's (even earlier in IT industry), the San Francisco-Oakland-San Jose area has surpassed other MSAs/CMSAs in both total number of patents, IT patents, and AI patents (seen in figures 2.1, 2.2, and 2.3); meanwhile, west coast cities lead with 61.7% in value and 39.5 % out of 130.9 billion dollars invested across 8,948 deals reported in 2018[1]. Indeed, the concentration of invention activities echos the focus on the local expansion of VC investment activities. Sorenson and Stuart [23] evaluates the concentration of venture capitalists, that is the information of contacts and connections is more likely to be shared within regions. Samila and Sorenson [20] documented their findings on areas that are VC-backed, or mentioned by them as in the "local venture capital community", has more substantial effectiveness with converting academic research into innovations with VC supply. Chen, Gompers, and Lerner (2010) [5] show that there exists the geographic concentration by venture capital firms and their targeted companies in San Francisco, Boston, and New York.

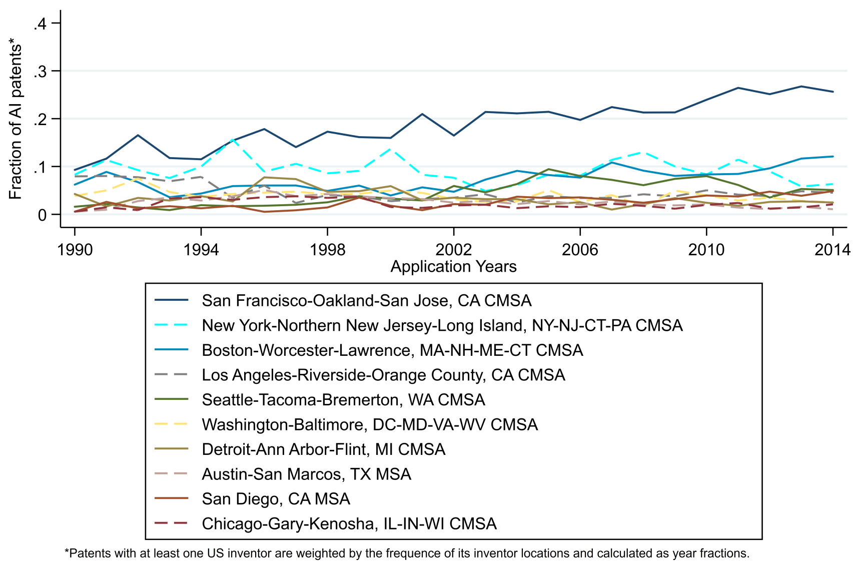

Figure 2.3: Fraction of AI Patents in Top 10 MSAs/CMSAs

[1]Sources: https://pitchbook.com/news/articles/18-charts-to-illustrate-us-vc-in-2018

2.2 Public and private funding: competition or cooperation?

While many studies examine economic outcomes or the market value of inventive actives [1] [8], there is a growing interest in how financial investments can fuel innovation reciprocally [2] [14] [20] [21]. Wallsten (2000) [24] attempts to estimate the effect of government grants on private innovation outcomes. Lach (2002) [15] extensively studies the rewards of R&D subsidizes using data on Israeli manufacturing firms in the 1990s. Additionally, mixed evidence is founded about whether public research fund could significantly increase the innovation outcomes [18] [24]. By contrast, studies on private investment reveal that there exist positive effects of financial investments on foresting adoption of technologies and establishment of firms [19] [20]. A recent work of Breznitz, Forman, and Wen (2018) [3] suggests that experience in a particular industry will increase the complementarity between VC financing and the introduction of new products for firms.

The comparison between public and private research funding also plays a major part in the discussion of the field. Few studies compare the efficiency of private and public grants [2] [8] [18]. However, one can still be challenged given the multidimensional heterogeneity of regions, products, and business types. Questions are also asked about if the government funding is a complement or substitute of private research funding. To provide answers, Samila and Sorenson (2010) [20] use a panel data on metropolitan areas to demonstrate the complementary effect between VC investment and technology commercialization outcomes; Muscio, Quaglione, and Vallanti (2013) [16] use Probit and Tobit panel data models to estimate on financial data and argue that government funding to universities complements with private research funding.

CHAPTER 3

DATA

This thesis assembled a panel data from 1990 to 2014 for 248 MSAs/CMSAs.[1] The panel data summary, as shown in table 3.1, contains eight variables in five categories: 1) annual census estimates of population; 2) federal support for research and development; 3) the count of patents, AI patents, and non-AI IT patents with at least one US inventors and weighted by the frequency of investor locations; 4) VC deal measurements including the number of deals and the disclosed amount achieved at each round of VC deal as with portfolio companies; 5) annual average return rates to limited partners (LP).[2]

Variables	Mean	Std. Dev.	N
Population (thousands)	914.3	2,088.4	6,200
Federal R&D Fund (millions)	75.5	215.9	6,200
Patents	330.1	1,193.7	6,200
AI Patents	1.6	8.2	6,200
Non-AI IT Patents	31.2	146.6	6,200
VC Deal Count	15.8	89.9	6,200
VC Deal Amount (disclosed, millions)	118.2	922.0	6,200
LP Return (%)	0.2	0.5	6,200

Notes: Variables are at the level of MSA/CMSA-year.

Table 3.1: Summary Statistics

To control for data quality, I set the time window as between 1990 and 2014[3]. First, there is usually a 3-5 year lag between the patent application date and grant date. Next, there is a change in classification system occurring between 2013 and 2015.[4] To avoid the pollution from the lag period and classification

[1] The definition of MSA/CMSA in this paper is the version released by U.S. Census Bureau on December 7, 1999 with revision on May 1, 2003. Even though the Census Bureau released an updated crosswalk for CBSA/MSA/PMSA, this thesis uses the 1999 definition only for logical consistency. According to the 1999 definition, there are overall 280 MSAs/CMSAs.

[2] Construction of LP return follows the approach from Samila and Sorenson (2010). [20] The construction and modification will be specified in later sections.

[3] All the dates associated with the variables is obtained in the calendar year.

[4] USPTO has been gradually moved from USPC to CPC to classify the US granted patents

change, I ended the time-frame when the total number of patents starts to drop dramatically (as shown in figure A.1). I mapped inventors to counties and metropolitan statistical areas (MSAs) and consolidated MSAs (CMSAs) if applicable. The reason for utilizing MSA/CMSA as the unit of aggregation follows the logic from Forman, Goldfarb, and Greenstein (2016) [9] and Samilla and Sorenson (2010) [20]. First, based on the assumption that VC funds are usually invested locally. Second, based on the fact that technological inventions have a propensity to agglomerate in urban regions. Consequently, the notion of metropolitan statistical areas suits the nature of this paper's analyses. Besides, I kept 248 MSAs/CMSAs (out of 280), which with at least VC investment data available. For MSAs/CMSAs with missing VC measurements, the assumption is that the relationship between patenting behavior and VC investment is unknown for those areas, and hence, I excluded them.

3.1 Data sources

The source of population data is census U.S. intercensal county population data from the National Bureau of Economic Research (NBER).[5] Federal RD fund data is extracted from the results of the Survey of Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions from the National Science Foundation (NSF).[6] Patents data is obtained from the PatentsView datasets managed by the U.S. Patent and Trademark Office (USPTO).[7] VC deal

starting on Jan 2013. (Source: https://www.uspto.gov/patents-application-process/patent-search/classification-standards-and-development) However, there is no official notice on the actual end date when the USPTO stopped issuing USPC codes for patents. From data accessed on PatentsView, I observed the change stopped on Jan, 2015.

[5]Source: https://www.nber.org/data/census-intercensal-county-population.html

[6]Source: https://ncsesdata.nsf.gov/ids/fss

[7]Source: http://www.patentsview.org/download/

data is downloaded through SDC Platinum by Thomson Reuters[8]. The annual average endowment return rates are available from the National Association of College and University Business Officers (NACUBO) historic endowment study data.[9] The next section will elaborate more on data consolidation process and variable generation.

3.2 Variable description

3.2.1 Dependent variables

Utility Patents (Patents). The original patents granted in the US are first filtered by patent type equals "Utility". The reason this thesis is using utility patents only is that utility patents are mainly on products, process, and machine, which is proper for the scope of technological invention. Under the selection of US granted utility patents, I first identified each patent's inventors addresses as domestic or foreign and then cross-walked the domestics addresses to MSA/CMSA code. Particularly, I geocoded the addresses of inventors using Google API and matched the geo-coordinates with FIPS through overlaying functions in ArcGIS; then, I matched with the MSA/CMSA code for each FIPS code under the 1999 definition [10]. Then, I dropped those patents with all foreign inventors. In addition, I dropped the patents before 1990 and after 2014. Finally, patents are counted (and logged) within each MSA/CMSA per year. In summary, the final selection of US granted patents are restricted to utility patents

[8] Date of access: June 05, 2019
[9] Source: https://www.nacubo.org/Research/2019/Historic-Endowment-Study-Data
[10] Source: https://www.census.gov/population/estimates/metro-city/99mfips.txt.

that are applied between 1990 and 2014 and with at least one US inventor[11]. To simplify the naming of this variable, I will refer to this selection as ”Patents”.

AI Patents. The selection of AI patents uses the patents selected by the criteria above and applies a definition of AI patent classification from Cockburn, Henderson, Stern (2018) [6]. AI patents are identified within the patent selection in two additional ways: 1) by using USPC classification[12] 706 (Data Processing-Artificial Intelligence) and 901 (Robots), and 2) by searching patent titles utilizing a list of keywords[13]. After taking the union of patents obtained from 1) and 2), I deleted the duplicates and counted (and logged) the number of AI patents within each MSA/CMSA per year.

Non-AI IT Patents (IT Patents). The selection of non-AI IT patents is also based on the utility patent selection. Following Forman, Goldfarb, and Greenstein (2016) [9], I identified the patents with USPC code under HJT subcategory 22 (Computer Hardware and Software) and 53 (Motor, Engines, and Parts) using a USPC-based patent subject classification developed by Hall, Jaffe and Trajtenberg (2001) [10]. I used these two subcategories because the two USPC classifications that are used to identify AI patents are indented under them, respectively. Therefore, patents with HJT 22 and 53 are supposed to be a broader selection that contains AI patents and the closest comparison of AI patents. To construct the comparison group only, I excluded the AI patents and counted (and logged) the rest of patents with HJT subcategory 22 and 53 within each MSA/CMSA per year. I will later abbreviate this selection of patents as ”IT Patents”.

[11] I also dropped patents with application date greater than grant date and with no FIPS code matched. Fortunately, those patents are only in a negligible fraction.

[12] U.S. Patent Classification System (USPC) is used by USPTO to categorize patents by their major and minor components.

[13] The list of keyword is available at the Appendix in Cockburn, Henderson, Stern (2018) [6]

3.2.2 Independent variables

VC Activities Measurements (VC Deal Count; VC Deal Amount). VC activities measurements are aggregated from raw data as with portfolio companies at the deal level. I first selected deals from 1990 to 2014 by round date and then kept only the deals where companies are at the stage (level 2) of *Seed, Early Stage, Later Stage, and Expansion.* The reason of including those stages before companies going to public or in market is that the instrument variable requires one to include only deals invested by VC funds with limited partners. [20] Using the zip-code of portfolio companies, I located each deal to MSAs/CMSAs and counted (and logged) the number of VC deals within each MSA/CMSA per year. As a second way to measure VC investment activities, I also aggregated (and logged) the amount disclosed for each VC deal within each MSA/CMSA per year. A special reminder about the aggregation method as compared with Samila and Sorenson's is that this thesis applied locations of portfolio companies instead of locations of VC fund. Essentially, VC funds will be used for investments by VC firms to portfolio companies. To be more precisely, the capital flow starts from the VC funds and goes to targeted portfolio companies. In Samila Sorensons paper [20], they argue that they are using the starting location (i.e. VC fund location) of the capital flow. This thesis, nevertheless, uses the destination of the same cash flow to locate the capital flow.

Federal R&D Fund. Federal R&D fund data is originally acquired at the institutional level with institutional names and state information. I selected the federal R&D fund from 1990 to 2014, geo-coded the addresses of inventors and matched the geo-coordinates with FIPS then MSA/CMSA code. The amount of federal R&D fund is finally counted (and logged) within each MSA/CMSA per year.

CHAPTER 4

EMPIRICAL STRATEGIES

4.1 Empirical model

$$\ln Pat_{i,t} = \alpha + \beta_1 \ln Pop_{i,t} + \beta_2 \ln RD_{i,t} + \beta_3 \ln VC_{i,t} + \beta_4 \ln VC_{i,t} \ln RD_{i,t} + \omega_t + \eta_i + \epsilon_{i,t} \tag{4.1}$$

$$\ln AIPat_{i,t} = \alpha + \beta_1 \ln Pop_{i,t} + \beta_2 \ln RD_{i,t} + \beta_3 \ln VC_{i,t} + \beta_4 \ln VC_{i,t} \ln RD_{i,t} + \omega_t + \eta_i + \epsilon_{i,t} \tag{4.2}$$

$$\ln ITPat_{i,t} = \alpha + \beta_1 \ln Pop_{i,t} + \beta_2 \ln RD_{i,t} + \beta_3 \ln VC_{i,t} + \beta_4 \ln VC_{i,t} \ln RD_{i,t} + \omega_t + \eta_i + \epsilon_{i,t}, \tag{4.3}$$

where i indexes the MSA and t represents year (calendar year). $Pat_{i,t}$, $AIPat_{i,t}$, and $ITPat_{i,t}$ are dependent variables used at different stages in each regression. $Pop_{i,t}$ controls for population size. $RD_{i,t}$ measures the federal R&D fund inflows and $VC_{i,t}$ measures the VC activities at the region level. $VC_{i,t}$ contains two variables, namely, $VCcount_{i,t}$ and $VCamount_{i,t}$. To capture the complementary effect between public research funding and VC investments, the model also includes the interaction between federal R&D fund and $VCcount_{i,t}$ and $VCamount_{i,t}$, respectively. ω_t denotes the year-specific fixed effects (year dummies); η_i represents the MSA/CMSA-specific fixed effect; while, $\epsilon_{i,t}$ means the error term. Here, the year-specific fixed effects ascribe to explain the change in national economy climate accross MSA/CMSA over time. By contrast, the MSA/CMSA-specified fixed effects attribute to all the characteristics within MSA/CMSAs.

4.2 Identification

4.2.1 Construction of instrumental variable

Through including MSA/CMSA-specific fixed effects, the model has identified the heterogeneity in different MSAs/CMSAs. However, concerns may also arise from unobserved factors that are correlated with the amount of local VC investment supply and other unobservables. As touched upon in Samila and Sorenson (2010) [20] and Samila and Sorenson (2011) [21], the return of VC investment to local investors produces different incentives for local VC investors. This thesis adopted their assumptions and construction of IV with a slight modification. The initial LP return consists of the annual endowment return rate and the number of limited partners (LP) who has invested in any fund within the desirable time-frame. By contrast, the number of investors I used to generate LP return is the number of LP investors at each VC deal. By changing the total number of LPs invested in funds a particular MSA/CMSA to the number of LPs participated in each deal, the construction of LP return was adjusted to the same level of aggregation as other VC measurements.

LP returns. To summarize it, the formula is given by:

$$\ln LPR_{i,t} = \sum_{s=t-1}^{t-3} ER_s \ \ln(1 + LPcount_{i,s}) \tag{4.4}$$

where i indexes the MSA and t represents year (calendar year). ER_s represents the annual average return rates to U.S. higher education endowments and affiliated foundations in %, and $LPcount_{i,s}$ is the number of LP investors for each VC deal.

The justification of using this IV on VC investment measures generally followed the assumption proposed by Samila and Sorenson (2010) [20]. First of all, LP returns are proved to be the primary inventive to limited partners since institutional investors often refer to an unchanged rate to allocate their assets across different asset classes. The fixed asset allocation rate, however, is usually determined by the historical return rates by asset class. [21] Second, the preference of institutional investors are not correlated with other variables in the equations. This propensity is described as ”home bias” by Samila and Sorenson (2010) in their paper. Third, ”home bias” existed in VC funds when they make investments on portfolio companies. The three assumptions overall guaranteed that the IV, LP return, influences VC investments without being correlated with omitted variables that will affect patenting.

4.2.2 MSA/CMSA selection and robustness checks

The selection of top 150 MSAs/CMSAs is driven by the fact that the panel is extremely unbalanced with missing VC investments measurements and public research fund for smaller MSAs/CMSAs. Apart from the analysis that analogous to Samila and Sorenson (2010) [20], the assumption from this thesis is that both inventive and VC investment activities are negligible in small MSAs/CMSAs. The reason of excluding small MSAs/CMSAs with little or none VC investments is that the relationship between local investment and patenting is likely to be different in these locations and that some of these locations may not have any inventive activity (or at least the type that shows up in patents).

To test the resilience of the main model, I constructed the robustness checks in five ways: i) changing the criteria of choosing top MSAs/CMSAs; ii) decreasing the number of top MSA included in main regression tables; iii) including all MSAs/CMAs in main regression tables; iv) altering the AI definition by using a different searching strategy; v) using a subset of patents selection which contains USPC code for all patents.[1] By changing the number of MSA/CMSAs included in the analysis, I sought to test if excluding more small MSA/CMSAs can affect the results and if the exclusion is crucial for improving the model. To determine if the AI definition imported is reasonably reliable, I employed another searching strategy when selecting AI patents using patent titles. Specifically, I allowed intervals or changing of orders for keywords when searching and consequently enlarged the AI patents selection. Eventually, I selected patents with only USPC codes, sorted out the AI and non-AI IT patents using the same strategies for the entire collection of patents. As one may recall, there exists an inconsistency in defining AI and non-AI IT patents. Two components that comprise the AI patents selection are patents with USPC subclass 706 and 901 and patents that contain keywords in their titles. Nevertheless, non-AI IT patents are only identified by USPC sub-classes fall into HJT class 22 and 53. It could be inferred, therefore, that some of the AI patents are not within the scope of IT patents since USPTO started issuing less USPC code to the patents granted after 2013. This complexity is better visualized by figures A.2 and A.3 in Chapter A of the Appendix. With a compromise of losing new patents and new AI patents without USPC code, I provided the subset of patents with USPC code to replicate the key analysis to avoid the definition inconsistency.

[1]Detailed results can be found in Chapter C of the Appendix.

CHAPTER 5

EMPIRICAL RESULTS

Following equations 4.1, 4.2, and 4.3, I conducted OLS regressions with fixed-effect estimates on utility patents, AI patents, and non-AI IT patents, respectively, for top 150 MSAs/CMSAs by total number of patents. To better present the results, the analysis will be segmented into four parts as table C.1, table C.2. Table 5.1 compares well with the main results from Samila and Sorenson (2010) [20]; whereas, tables C.1 and C.2 are analyses to show the difference in complementary effects on AI patents and non-AI IT patents. Following the construction of IV, I instrumented LP return on VC deal counts in main tables C.1 and C.2, respectively, and presented the results in tables 5.2.

5.1 Fixed effects estimation results

From Table 5.1, the OLS regression with fixed-effect estimators on VC deal count and amount in columns (1) and (3), respectively, indicates that both public research fund and VC activities accelerate the rates of patenting after controlling for variations across and within MSAs/CMSAs. With population and federal R&D fund fixed within one MSA/CMSA, the elasticity of the counts of VC deals with respect to the count of patents is 0.0484. In other words, if one MSA/CMSA has its count of VC deal doubled, there will be 11 more patents applied on average within a year ($0.0484 \times \ln 2 \times 330.1$ (mean of Patents) $= 11.07$). With population and federal R&D fund fixed within one MSA/CMSA, the elasticity of the aggregated amount of VC deal in millions with respect to the count of patents is 0.0244. Similarly, if one MSA/CMSA has its size of aggregated VC deal amount doubled, there would be 6 more patents applied on average within a year ($.0244 \times \ln 2 \times 330.1$ (mean of Patents) = 5.58). In columns (2) and (4), the

VARIABLES	(1) Patents	(2) Patents	(3) Patents	(4) Patents
Population (t-1)	1.111***	1.090***	1.097***	1.081***
	(0.129)	(0.128)	(0.130)	(0.128)
Federal R&D Fund (t-1)	0.0802***	0.0474**	0.0807***	0.0584***
	(0.0210)	(0.0211)	(0.0213)	(0.0208)
VC Deal Count (t)	0.0484***	-0.0260		
	(0.0129)	(0.0185)		
VC Deal Count (t)				
Federal R&D Fund (t-1)		0.0267***		
		(0.00614)		
VC Deal Amount (t)			0.0244***	-0.00735
			(0.00570)	(0.00886)
VC Deal Amount (t)				
Federal R&D Fund (t-1)				0.0124***
				(0.00321)
Observations	5,952	5,952	5,952	5,952
R-squared	0.215	0.225	0.215	0.221
Number of MSA/CMSA	248	248	248	248
Time Period	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table 5.1: Fixed-effect estimation results for all US utility patents on all MSAs/CMSAs

coefficients on interaction terms are both positive and significant. With an additional calculation using results from column (2), the elasticity of the counts of VC deals with respect to the count of patents arrives at -0.136 ($-0.0260 + 0.0267 \times \ln 0.016 = -0.136$) when Federal R&D Fund is low (p25 of Federal R&D Fund in millions) and arrives at 0.080 ($-0.0260 + 0.0267 \times \ln 53.5335 = 0.080$) when it is high (p75 of Federal R&D Fund in millions). With an additional calculation using results from column (4), the elasticity of the counts of VC deals with respect to the count of patents arrives at -0.059 ($-0.00735 + 0.0124 \times \ln 0.016 = -0.059$) when Federal R&D Fund is low and arrives at 0.042 ($-0.00735 + 0.0124 \times \ln 53.5335 = 0.042$) when it is high. Therefore, results from table 5.1 confirm that there exists a positive relationship between VC investment activities and rates of patenting and that there also exists the complementary relationship between VC investment activities and federal R&D funding as suggested by Samila and Sorenson (2010) [20] and Muscio, Quaglione, and Vallanti (2013) [16]. Additionally, results here show that increasing VC deals could only increase the rates of patenting when the presence of public research funding is high.[1]

Table 5.2 and table 5.3 are two OLS regressions with fixed effects estimates for top 150 MSAs/CMSAs by total number of patents using two VC investment measurements. Please remind that the number of observations drops because only the top 150 MSAs/CMSAs are included in the regressions this time. As can be seen from columns (1), (3), and (5) in the two tables, the coefficients on Federal R&D Fund lose significance on the rates of IT patenting and remain significant for AI and the entire patents selection. Whereas, all the coefficients on VC activities measurement remain positive and significant. The comparison suggests that the public research funding seems not to significantly influence IT patenting behavior; while, change in VC investment activities remains to affect the growth of AI and IT patents. On condition of within top 150 MSAs/CMSAs, the elasticity of Federal R&D Fund with respect to the count of patents is the highest for AI patents (0.0544) with comparison of 0.0543 for non-AI IT patents and 0.0397 for all patents; the elasticity of the counts of VC deals with respect to the count of patents is the highest for AI patents (0.0609) with comparison of 0.0514 for non-AI IT patents and 0.0461 for all patents. Similarly, the elasticity of Federal R&D Fund with respect to the count of patents is the highest for AI patents (0.0551) with comparison of 0.0538 for non-AI IT patents and 0.0420 for all patents; the elasticity of the aggregated amount of VC deals with respect to

[1] Precisely, the Federal R&D Fund amount in millions need be greater than 0.102 when using VC Deal Count as VC investment measurement and 0.042 when using VC Deal Amount.

the count of patents is also the highest for AI patents (0.0296) with comparison of 0.0191 for non-AI IT patents and 0.0227 for patents.

VARIABLES	(1) Patents	(2) Patents	(3) AI Patents	(4) AI Patents	(5) IT Patents	(6) IT Patents
Population (t-1)	1.085***	1.077***	0.345*	0.328*	1.195***	1.181***
	(0.157)	(0.156)	(0.182)	(0.168)	(0.226)	(0.228)
Federal R&D Fund (t-1)	0.0543**	0.0125	0.0544*	-0.0286	0.0397	-0.0288
	(0.0252)	(0.0257)	(0.0281)	(0.0268)	(0.0423)	(0.0442)
VC Deal Count (t)	0.0348**	-0.0404*	0.0646***	-0.0848***	0.0482**	-0.0752**
	(0.0147)	(0.0223)	(0.0211)	(0.0296)	(0.0228)	(0.0360)
VC Deal Count (t)						
Federal R&D Fund (t-1)		0.0246***		0.0488***		0.0403***
		(0.00692)		(0.00983)		(0.0104)
Observations	3,600	3,600	3,600	3,600	3,600	3,600
R-squared	0.385	0.398	0.107	0.135	0.488	0.495
Number of MSA/CMSA	150	150	150	150	150	150
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table 5.2: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC count (top 150 MSAs/CMSAs selected by total patent number)

VARIABLES	(1) Patents	(2) Patents	(3) AI Patents	(4) AI Patents	(5) IT Patents	(6) IT Patents
Population (t-1)	1.071***	1.064***	0.323*	0.309*	1.184***	1.171***
	(0.158)	(0.156)	(0.185)	(0.172)	(0.225)	(0.227)
Federal R&D Fund (t-1)	0.0538**	0.0268	0.0551**	-0.00327	0.0420	-0.0165
	(0.0254)	(0.0248)	(0.0279)	(0.0258)	(0.0431)	(0.0429)
VC Deal Amount (t)	0.0180***	-0.0128	0.0281***	-0.0384***	0.0156	-0.0511***
	(0.00608)	(0.00994)	(0.00974)	(0.0142)	(0.0104)	(0.0188)
VC Deal Amount (t)						
Federal R&D Fund (t-1)		0.0109***		0.0235***		0.0236***
		(0.00358)		(0.00534)		(0.00545)
Observations	3,600	3,600	3,600	3,600	3,600	3,600
R-squared	0.385	0.394	0.106	0.127	0.487	0.495
Number of MSA/CMSA	150	150	150	150	150	150
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table 5.3: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC amount (top 150 MSAs/CMSAs selected by total patent number)

From columns (2), (4), and (6) in tables 5.2 and 5.3, it is also suggested that the coefficients on the interaction terms between Federal R&D Fund and VC investment activities on three selections of patents remain positive and significant. Specifically, from columns (2), (4), and (6) in table 5.2, the elasticity of the counts of VC deals with respect to the count of patents arrives at -0.056 ($-0.0404 + 0.0246 \times \ln 0.522 = -0.056$) when Federal R&D Fund is low (p25 of Federal R&D Fund in millions for Top 150 MSAs/CMSAs) and arrives at 0.071 ($-0.0404+0.0246\times\ln 91.63701 = 0.071$) when it is high (p75 of Federal R&D Fund in millions for Top 150 MSAs/CMSAs). The elasticity of the counts of VC deals with respect to the count of AI patents arrives at -0.117 when Federal R&D Fund is low and arrives at 0.136 when it is high. The elasticity of the counts of VC deals with respect to the count of IT patents arrives at -0.101 when Federal R&D Fund is low and arrives at 0.107 when it is high. From columns (2), (4), and (6) in table 5.3, the elasticity of the aggregated amount of VC deals with respect to the count of patents arrives at -0.020 when Federal R&D Fund is low (p25 of Federal R&D Fund in millions for Top 150 MSAs/CMSAs) and arrives at 0.036 when it is high (p75 of Federal R&D Fund in millions for Top 150 MSAs/CMSAs). The elasticity of the counts of VC deals with respect to the aggregated amount of AI patents arrives at -0.054 when Federal R&D Fund is low and arrives at 0.068 when it is high. The elasticity of the aggregated amount of VC deals with respect to the count of IT patents arrives at -0.066 when Federal R&D Fund is low and arrives at 0.056 when it is high. Comparison of AI with IT patents provides evidence of the positive effect of VC investment activities and the complementary relationship between VC investment activities and Federal R&D Fund on AI and IT patenting. The elasticity of two measurements of VC investment activities and Federal R&D Fund on AI patents are all the highest. As mentioned in table 5.1, VC investments also only increase the rates of AI and IT patenting when the presence of Federal R&D Fund is high. Moreover, the elasticity of two measurements of VC investment activities on AI patents is the highest given a high presence of Federal R&D Fund. Finally, results after robustness checks generally show consistency with the main regression tables 5.2 and 5.3. Nevertheless, the significance of Federal R&D Fund appears occasionally on the rates of AI patenting when altering the criteria of selecting top MSA/CMSA by annual average population, changing the number of MSA/CMSA selected, and applying a flexible search of AI patents.

5.2 Instrumental variable estimation results

While the fixed effects estimation help to identify the variation across and within MSAs/CMSAs, there might still be unobserved county-level factors that motivated one to apply the IV estimates. A rough assessment for the four tables reveals that VC deal count has a consistently larger elasticity on the rates of patenting for three patent selections. Therefore, I only applied the IV estimates on VC count. As seen from the first stage results, the coefficient on IV constructed is significant when predicting the VC deal count. As seen in columns (2), (4), and (6), both coefficients on Federal R&D Fund and VC deal count remain positive for patents and AI patents but not for IT patents. However, the significance level loses for most of the coefficients. As seen in columns (3), (5), and (7), the significance of all interaction terms disappeared except for IT patents. After an additional calculation, the elasticity of the aggregated amount of VC deals with respect to the count of patents arrives at 0.112 when Federal R&D Fund is low (p25 of Federal R&D Fund in millions for Top 150

MSAs/CMSAs) and arrives at 0.110 when it is high (p75 of Federal R&D Fund in millions for Top 150 MSAs/CMSAs). The elasticity of the counts of VC deals with respect to the aggregated amount of AI patents arrives at 0.179 when Federal R&D Fund is low and arrives at 0.021 when it is high. The elasticity of the aggregated amount of VC deals with respect to the count of IT patents arrives at -0.022 when Federal R&D Fund is low and arrives at 0.103 when it is high. In other words, both the positive effect of VC investments and complementary effect are not robust to OLS results after importing IV estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	First Stage	Patents	Patents	AI Patents	AI Patents	IT Patents	IT Patents
LP Returns	0.313***						
	(0.0353)						
Population (t-1)	0.100	1.075***	1.075***	0.348***	0.355***	1.190***	1.184***
	(0.123)	(0.0616)	(0.0620)	(0.0982)	(0.101)	(0.133)	(0.133)
Federal R&D Fund (t-1)	0.166***	0.0425***	0.0430	0.0583**	0.106	0.0335	-0.00430
	(0.0271)	(0.0152)	(0.0330)	(0.0238)	(0.0752)	(0.0301)	(0.0569)
VC Deal Count (t)		0.111**	0.112	0.0395	0.159	0.0881	-0.00620
		(0.0524)	(0.0737)	(0.0972)	(0.132)	(0.0996)	(0.141)
VC Deal Count (t) Federal R&D Fund (t-1)			-0.000358		-0.0306		0.0241
			(0.0173)		(0.0405)		(0.0300)
Observations	3,600	3,600	3,600	3,600	3,600	3,600	3,600
R-squared		0.369	0.368	0.106	0.062	0.487	0.494
Number of MSA/CMSA	150	150	150	150	150	150	150
F-test (IV excl.)		78.70	43.97	78.70	43.97	78.70	20.12
F-test (IV intersection excl.)		NA	20.12	NA	20.12	NA	43.97
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table 5.4: IV estimation results for all US utility patents, AI patents, and non-AI IT patents on VC count (top 150 MSAs/CMSAs selected by total patent number)

CHAPTER 6

CONCLUSIONS

The strategy and framework devised in this thesis mainly contribute as a preliminary attempt to document the recent regional patterns of us utility patents and AI patents. Meanwhile, the thesis also intends to evaluate the complementary effect between public and private funding on accelerating inventive activities of cutting-edge technology like AI. To develop the analysis in depth, I also constructed a comparison group containing non-AI but IT-related patents to investigate more on the difference in complementary effects between the frontier and traditional technologies.

Using the panel data from 1990 to 2014 on top 150 metropolitan statistical areas (MSAs) and consolidated metropolitan statistical areas (CMSAs) selected by population and total patent number, I applied lag-structured OLS regressions on US utility patents, US AI patents, and US non-AI IT patents, respectively, with Fixed-Effect estimates and Instrumental estimates. Aligning with the previous findings from Samila and Sorenson (2010) [20], the fixed effects results from table 5.1 revealed that i) VC investments played a decisive role in stimulating patenting behavior; ii) the participation of VC investment would much encourage the rate of patents with the presence of high public funding amount. The comparison between three patent selections from tables 5.2 and 5.3 indicated that the federal R&D support is less crucial on encouraging the growth of traditional IT patents with the comparison of AI patents. By contrast, there exists a significant and positive role of VC investments and the significant and positive complementary effect between VC investments and federal R&D funds for both AI and IT patents. Despite the difference in significance level, both Federal R&D Fund and VC investments have the highest elasticity on the rates of AI patenting. This fact implied that the development of frontier technologies such as AI depends on both public and private financial support; while, traditional technological development may much rely on private financial support. A possible explanation is that new technologies are more likely to exist as start-ups and enterprises, which often have greater demands for financial aids.

After robustness checks, results remained robust to main OLS regressions with mixed evidence on if federal R&D fund played a significant role in stimulating AI patenting. After considering the heterogeneity of VC investments, I constructed the instrument variable (IV) named LP returns, as suggested by Samila and Sorenson (2011) [21]. Not only the significance level disappear on VC measurements and interaction terms, but it is also implied that the complementary effect disappears for patents and AI patents. A takeaway from the IV regressions is that the results with IV estimates failed to have a strong concordance with findings from Samila and Sorenson (2010) [20]. IV results might have weaker statistical significance and are not always consistent with the results from OLS. Due to the time constraints, unfortunately, it has not been feasible to provide explanations to these questions. Indeed, care should be taken to improve IV construction in this thesis.

CHAPTER 7

DISCUSSIONS

Back to the original study of the research questions, findings from this thesis supported previous findings on the positive relationship between VC investments and growth in patenting and extended this conclusion with evidence on AI patenting. Additional results from this study suggested that, however, federal R&D fund doesn't always significantly influence traditional IT patenting behaviors. After adding interactions between VC investment measurements and federal R&D fund, I found complementary effects existed on overall rates of patenting, AI patenting, and IT patenting. However, results from this thesis show that VC investments could only increase the rates of patenting when there is a high presence of Federal R&D Funding. Additionally, VC investment has the highest elasticity on the rates of AI patenting compared with the other two selections. Although with an intention to improve the estimations, both the positive relationship between VC investments and rates of patents and the complementary effect lost significance after instrumented by the annual average return to limited partners.

During the analysis, one of the difficulties this thesis has encountered is how to carefully choose a consistent definition of utility patents, AI patents, and non-AI IT patents. As one may point out, the time cut-off this thesis applied can only avoid the lag between application and grant date for utility patents and AI patents. The obstacle of making three perfect patent selections is largely caused by the change form USPC to CPC system. First of all, the end date when USPTO stopped issuing the USPC code is given by patent granted date instead of application date (not to mention that there was no an official end date when this change ended). The nature of this study, however, requires one to use patent application date so that the time lag between application and granted date will not pollute the lag structure of panel data. Secondly, although there now exists several statistical crosswalks that could map CPC back to USPC, the actual mapping presents enormous inconsistencies in the ways that CPC and USPC define their scope of classes and sub-classes. Usually, there will be several CPC matched with per USPC code (even after we only take the first classification per patent). In other words, it's almost unachievable to assign a unique CPC code per USPC code. Since both the definition of non-AI and part of the AI patents are developed based on USPC code, the ambiguity of crosswalk makes it hard to identify new patents that is AI or non-AI IT patents. Consequently, one should be extremely careful when including the most recent patents in both AI and non-AI selection using the definitions mentioned. However, the AI industry itself has witnessed its most rapid growth in very current time. The failure of including the most recent AI patent in the analysis might yield another "lag", which is the discrepancy between the results from this study and the latest trends in AI industrial development. In this thesis, however, I decided to end the time-frame early enough to avoid the crosswalk for most recent AI and IT patents. Ideally, a future crosswalk between USPC and CPC is planned to solve this dilemma by updating the definitions using CPC code for both AI and non-AI IT patents.

While the dilemma presented for the patent data used in this thesis, noteworthy concerns have been gained about the VC data. As discussed earlier in the conclusion section, the regression with IV estimates shows a major difference from previous findings. This problem motivated me to revisit the VC data. According to the summary statistics, the VC investments count from this thesis is generally comparable with previous findings from Samila and Sorenson [20]. However, the amount of VC investments are higher than their outcomes. The difference might come from: i) measurement error/biases of VC fund performance data as with portfolio companies [13]; or ii)different items for aggregation. It should, however, be noted that several funds usually participate in each VC investment round, and the amount a portfolio company received per round can be higher than the amount each VC fund invested per round. Nevertheless, one of the assumptions from this thesis is that the location of VC funds should be identical to (except across the MSA/CMSA border) the location of portfolio companies it invested at the MSA/CMSA level. The difference in amount should not influence the regression results theoretically. Future data quality checks should, therefore, be produced by aggregating VC deals using the same aggregation entity as Samila and Sorenson.

APPENDIX A

DESCRIPTIVE STATISTICS OF PATENT DATA

A.1 Utility patents

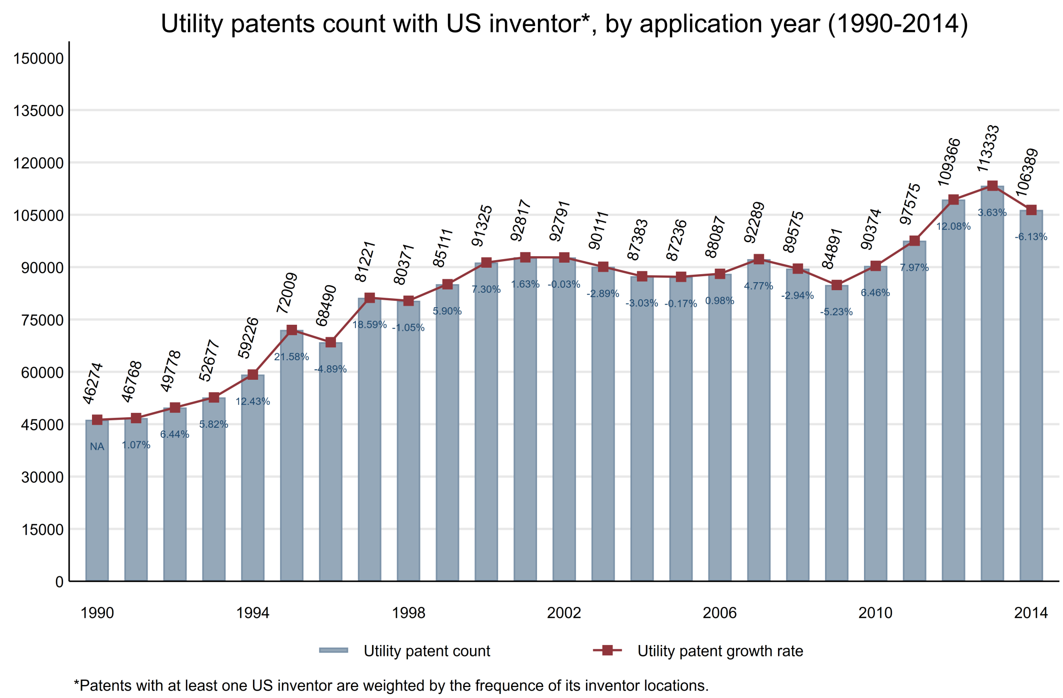

Figure A.1: Utility patents count with at least one US inventor

A.2 Non-AI IT patents

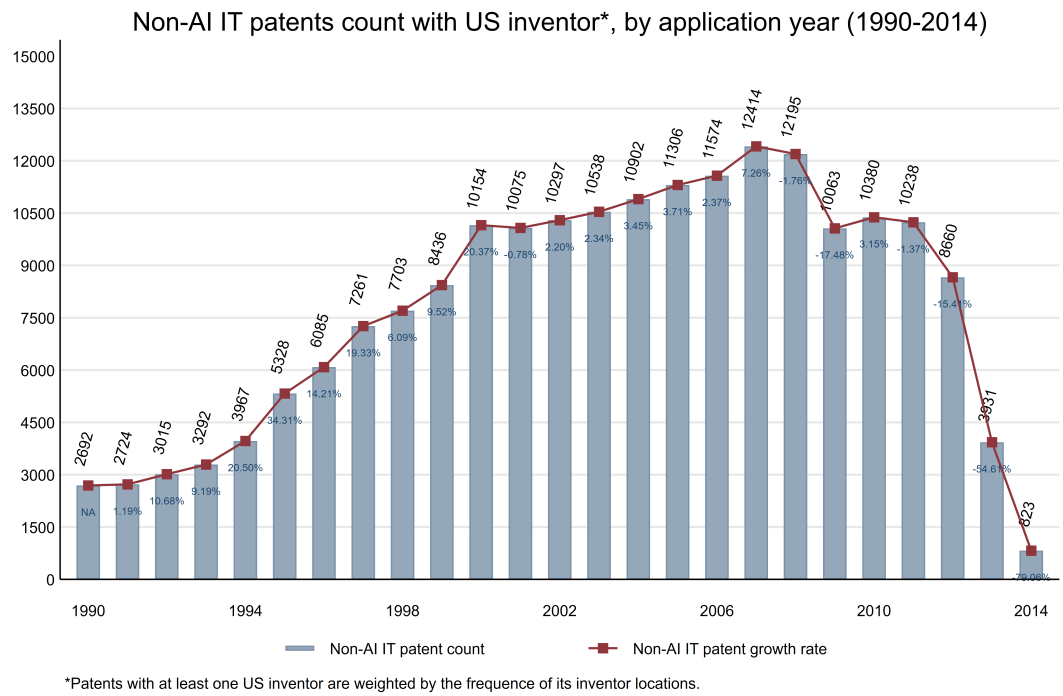

Figure A.2: Non-AI IT patents count with at least one US inventor

A.3 AI patents

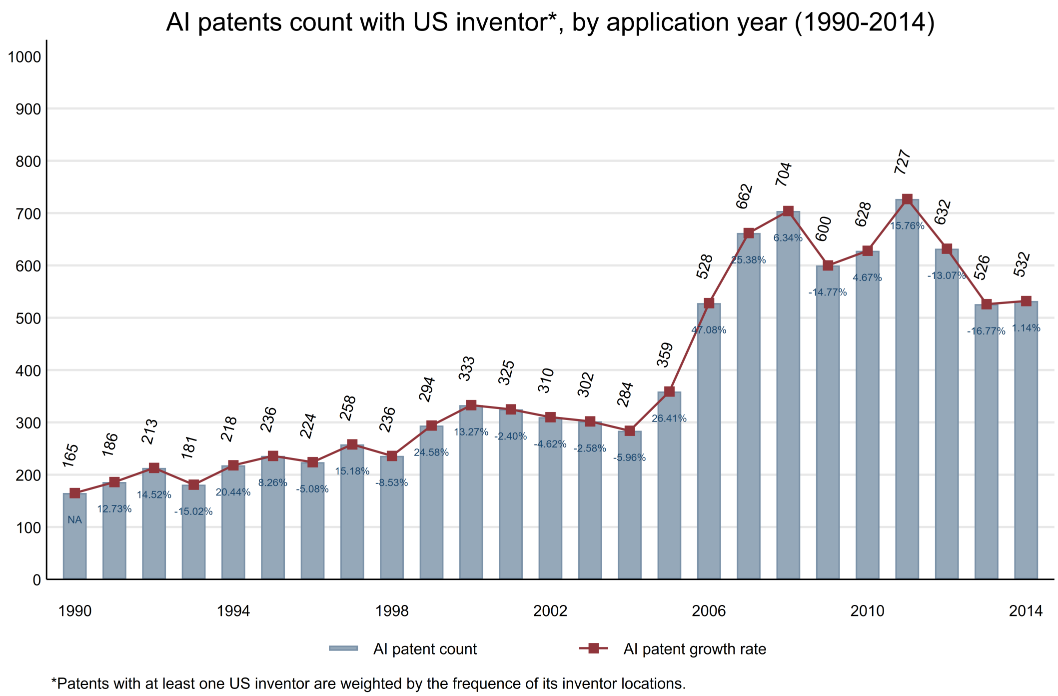

Figure A.3: AI patents count with at least one US inventor

APPENDIX B

DESCRIPTIVE STATISTICS OF VENTURE CAPITAL DATA

B.1 Venture capital investment count

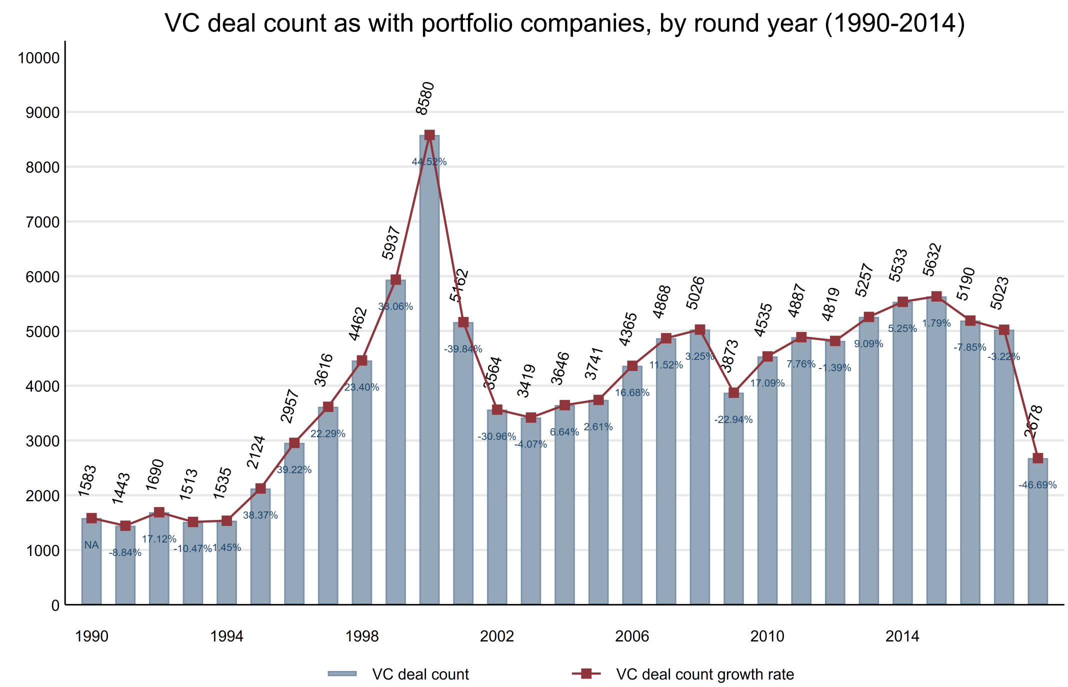

Figure B.1: VC deal count as with portfolio companies

B.2 Venture capital investment amount

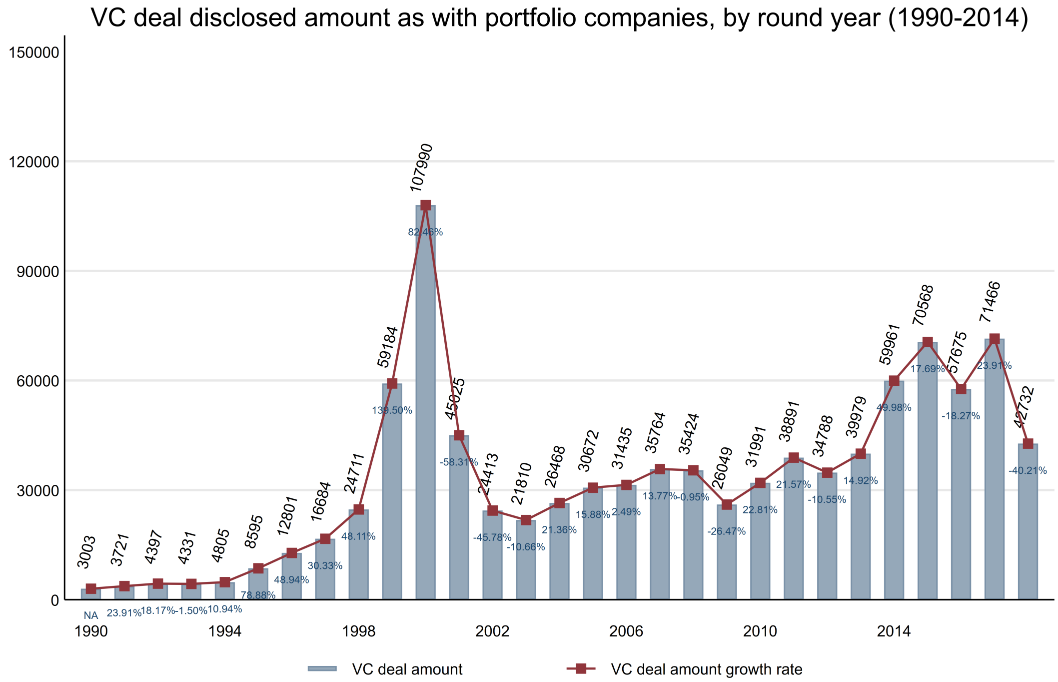

Figure B.2: VC deal disclosed amount as with portfolio companies

APPENDIX C

ROBUSTNESS CHECKS

C.1 Top 150 MSAs/CMSAs by annual average population

VARIABLES	(1) Patents	(2) Patents	(3) AI Patents	(4) AI Patents	(5) IT Patents	(6) IT Patents
Population (t-1)	1.260***	1.247***	0.343*	0.316*	1.248***	1.224***
	(0.145)	(0.142)	(0.180)	(0.163)	(0.221)	(0.222)
Federal R&D Fund (t-1)	0.0611***	0.0237	0.0356	-0.0404*	0.0298	-0.0391
	(0.0228)	(0.0237)	(0.0229)	(0.0221)	(0.0390)	(0.0388)
VC Deal Count (t)	0.0461***	-0.0291	0.0609***	-0.0917***	0.0514**	-0.0869**
	(0.0151)	(0.0226)	(0.0208)	(0.0274)	(0.0222)	(0.0347)
VC Deal Count (t) Federal R&D Fund (t-1)		0.0259***		0.0526***		0.0477***
		(0.00700)		(0.00995)		(0.0104)
Observations	3,600	3,600	3,600	3,600	3,600	3,600
R-squared	0.339	0.354	0.100	0.136	0.430	0.442
Number of MSA/CMSA	150	150	150	150	150	150
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table C.1: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC count (top 150 MSAs/CMSAs selected by annual average population)

VARIABLES	(1) Patents	(2) Patents	(3) AI Patents	(4) AI Patents	(5) IT Patents	(6) IT Patents
Population (t-1)	1.242***	1.232***	0.320*	0.298*	1.236***	1.214***
	(0.146)	(0.142)	(0.182)	(0.167)	(0.221)	(0.219)
Federal R&D Fund (t-1)	0.0606***	0.0355	0.0349	-0.0198	0.0308	-0.0261
	(0.0231)	(0.0230)	(0.0227)	(0.0209)	(0.0396)	(0.0380)
VC Deal Amount (t)	0.0227***	-0.00844	0.0296***	-0.0383***	0.0191*	-0.0515***
	(0.00643)	(0.0101)	(0.00958)	(0.0129)	(0.00991)	(0.0174)
VC Deal Amount (t)						
Federal R&D Fund (t-1)		0.0117***		0.0256***		0.0266***
		(0.00370)		(0.00538)		(0.00543)
Observations	3,600	3,600	3,600	3,600	3,600	3,600
R-squared	0.339	0.349	0.100	0.128	0.429	0.441
Number of MSA/CMSA	150	150	150	150	150	150
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table C.2: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC amount (top 150 MSAs/CMSAs selected by annual average population)

C.2 Alternative selection of top 100 MSAs/CMSAs by total patent number

VARIABLES	(1) Patents	(2) Patents	(3) AI Patents	(4) AI Patents	(5) IT Patents	(6) IT Patents
Population (t-1)	1.335***	1.333***	0.404	0.400*	1.655***	1.653***
	(0.178)	(0.179)	(0.252)	(0.237)	(0.265)	(0.269)
Federal R&D Fund (t-1)	0.0181	-0.0266	0.0308	-0.0715	0.0218	-0.0484
	(0.0370)	(0.0392)	(0.0444)	(0.0467)	(0.0515)	(0.0601)
VC Deal Count (t)	0.0316*	-0.0490	0.0655**	-0.119**	0.0420	-0.0847*
	(0.0179)	(0.0336)	(0.0282)	(0.0455)	(0.0273)	(0.0493)
VC Deal Count (t)						
Federal R&D Fund (t-1)		0.0236***		0.0539***		0.0370***
		(0.00873)		(0.0123)		(0.0125)
Observations	2,400	2,400	2,400	2,400	2,400	2,400
R-squared	0.473	0.485	0.132	0.157	0.637	0.643
Number of MSA/CMSA	100	100	100	100	100	100
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table C.3: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC count (top 100 MSAs/CMSAs selected by total patent number)

VARIABLES	(1) Patents	(2) Patents	(3) AI Patents	(4) AI Patents	(5) IT Patents	(6) IT Patents
Population (t-1)	1.324***	1.325***	0.381	0.383	1.646***	1.648***
	(0.179)	(0.178)	(0.258)	(0.240)	(0.265)	(0.265)
Federal R&D Fund (t-1)	0.0184	-0.00980	0.0311	-0.0432	0.0238	-0.0307
	(0.0372)	(0.0374)	(0.0441)	(0.0449)	(0.0526)	(0.0576)
VC Deal Amount (t)	0.0128*	-0.0208	0.0277**	-0.0607***	0.0130	-0.0519**
	(0.00750)	(0.0152)	(0.0131)	(0.0218)	(0.0132)	(0.0246)
VC Deal Amount (t)						
Federal R&D Fund (t-1)		0.0104**		0.0275***		0.0201***
		(0.00460)		(0.00655)		(0.00668)
Observations	2,400	2,400	2,400	2,400	2,400	2,400
R-squared	0.472	0.479	0.130	0.151	0.637	0.642
Number of MSA/CMSA	100	100	100	100	100	100
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table C.4: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC amount (top 100 MSAs/CMSAs selected by total patent number)

C.3 Selection of MSAs/CMSAs

VARIABLES	(1) Patents	(2) Patents	(3) AI Patents	(4) AI Patents	(5) IT Patents	(6) IT Patents
Population (t-1)	1.111***	1.090***	0.319**	0.280**	1.021***	0.985***
	(0.129)	(0.128)	(0.128)	(0.116)	(0.170)	(0.172)
Federal R&D Fund (t-1)	0.0802***	0.0474**	0.0479***	-0.0115	0.0478	-0.00722
	(0.0210)	(0.0211)	(0.0184)	(0.0155)	(0.0303)	(0.0295)
VC Deal Count (t)	0.0484***	-0.0260	0.0569***	-0.0781***	0.0618***	-0.0631**
	(0.0129)	(0.0185)	(0.0163)	(0.0218)	(0.0195)	(0.0275)
VC Deal Count (t)						
Federal R&D Fund (t-1)		0.0267***		0.0486***		0.0449***
		(0.00614)		(0.00864)		(0.00887)
Observations	5,952	5,952	5,952	5,952	5,952	5,952
R-squared	0.215	0.225	0.074	0.110	0.316	0.326
Number of MSA/CMSA	248	248	248	248	248	248
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table C.5: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC count (all MSA/CMSA)

C.4 Alternative definition of AI patents on top 150 MSAs/CMSAs by total patent number

VARIABLES	(1) Patents	(2) Patents	(3) AI Patents	(4) AI Patents	(5) IT Patents	(6) IT Patents
Population (t-1)	1.085***	1.077***	0.390**	0.373**	1.195***	1.181***
	(0.157)	(0.156)	(0.187)	(0.170)	(0.225)	(0.228)
Federal R&D Fund (t-1)	0.0543**	0.0125	0.0668**	-0.0200	0.0389	-0.0298
	(0.0252)	(0.0257)	(0.0278)	(0.0261)	(0.0421)	(0.0440)
VC Deal Count (t)	0.0348**	-0.0404*	0.0728***	-0.0834***	0.0480**	-0.0756**
	(0.0147)	(0.0223)	(0.0209)	(0.0291)	(0.0228)	(0.0360)
VC Deal Count (t)						
Federal R&D Fund (t-1)		0.0246***		0.0510***		0.0404***
		(0.00692)		(0.00960)		(0.0103)
Observations	3,600	3,600	3,600	3,600	3,600	3,600
R-squared	0.385	0.398	0.117	0.147	0.488	0.496
Number of MSA/CMSA	150	150	150	150	150	150
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table C.6: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC count (top 150 MSAs/CMSAs selected by total patent number)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Patents	Patents	AI Patents	AI Patents	IT Patents	IT Patents
Population (t-1)	1.071***	1.064***	0.367*	0.352**	1.184***	1.171***
	(0.158)	(0.156)	(0.190)	(0.175)	(0.225)	(0.227)
Federal R&D Fund (t-1)	0.0538**	0.0268	0.0682**	0.00549	0.0412	-0.0172
	(0.0254)	(0.0248)	(0.0276)	(0.0251)	(0.0429)	(0.0427)
VC Deal Amount (t)	0.0180***	-0.0128	0.0300***	-0.0415***	0.0155	-0.0511***
	(0.00608)	(0.00994)	(0.00958)	(0.0141)	(0.0104)	(0.0188)
VC Deal Amount (t)						
Federal R&D Fund (t-1)		0.0109***		0.0253***		0.0236***
		(0.00358)		(0.00520)		(0.00545)
Observations	3,600	3,600	3,600	3,600	3,600	3,600
R-squared	0.385	0.394	0.114	0.138	0.488	0.496
Number of MSAs/CMSAs/CMSA	150	150	150	150	150	150
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
Robust standard errors in parentheses.

Table C.7: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC amount (top 150 MSAs/CMSAs selected by total patent number)

C.5 Alternative selection of patents with USPC code on top 150 MSAs/CMSAs by total patent number

VARIABLES	(1) Patents	(2) Patents	(3) AI Patents	(4) AI Patents	(5) IT Patents	(6) IT Patents
Population (t-1)	1.138***	1.131***	0.374**	0.364**	1.195***	1.181***
	(0.156)	(0.156)	(0.147)	(0.143)	(0.225)	(0.228)
Federal R&D Fund (t-1)	0.0523**	0.0144	0.00864	-0.0432	0.0389	-0.0298
	(0.0257)	(0.0266)	(0.0253)	(0.0264)	(0.0421)	(0.0440)
VC Deal Count (t)	0.0243	-0.0440*	0.0517***	-0.0416	0.0480**	-0.0756**
	(0.0155)	(0.0224)	(0.0183)	(0.0258)	(0.0228)	(0.0360)
VC Deal Count (t)						
Federal R&D Fund (t-1)		0.0223***		0.0305***		0.0404***
		(0.00689)		(0.00897)		(0.0103)
Observations	3,600	3,600	3,600	3,600	3,600	3,600
R-squared	0.843	0.845	0.141	0.151	0.488	0.496
Number of MSA/CMSA	150	150	150	150	150	150
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table C.8: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC count (top 150 MSAs/CMSAs selected by total patent number)

VARIABLES	(1) Patents	(2) Patents	(3) AI Patents	(4) AI Patents	(5) IT Patents	(6) IT Patents
Population (t-1)	1.124***	1.118***	0.363**	0.353**	1.184***	1.171***
	(0.156)	(0.156)	(0.150)	(0.145)	(0.225)	(0.227)
Federal R&D Fund (t-1)	0.0507*	0.0251	0.0111	-0.0291	0.0412	-0.0172
	(0.0257)	(0.0256)	(0.0251)	(0.0252)	(0.0429)	(0.0427)
VC Deal Amount (t)	0.0162**	-0.0130	0.0169**	-0.0289**	0.0155	-0.0511***
	(0.00642)	(0.0101)	(0.00840)	(0.0129)	(0.0104)	(0.0188)
VC Deal Amount (t)						
Federal R&D Fund (t-1)		0.0103***		0.0162***		0.0236***
		(0.00355)		(0.00459)		(0.00545)
Observations	3,600	3,600	3,600	3,600	3,600	3,600
R-squared	0.843	0.845	0.138	0.148	0.488	0.496
Number of MSA/CMSA	150	150	150	150	150	150
Time Period	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014	1990-2014
Year Dummies	YES	YES	YES	YES	YES	YES
MSA/CMSA Fixed Effects	YES	YES	YES	YES	YES	YES

Notes: OLS regression, clustered by MSA/CMSA; *** p<0.01, ** p<0.05, * p<0.1.
Robust standard errors in parentheses.

Table C.9: Fixed-effect estimation results for all US utility patents, AI patents, and non-AI IT patents on VC amount (top 150 MSAs/CMSAs selected by total patent number)

BIBLIOGRAPHY

[1] James D. Adams. Fundamental Stocks of Knowledge and Productivity Growth. *Journal of Political Economy*, 98(4):673–702, 1990.

[2] Catherine Beaudry and Sedki Allaoui. Impact of public and private research funding on scientific production: The case of nanotechnology. *Research Policy*, 41(9):1589–1606, 2012.

[3] Dan Breznitz, Chris Forman, and Wen Wen. The Role of Venture Capital in the Formation of a New Technological Ecosystem: Evidence from the Cloud. *MIS Quartlery*, 42(4):1–28, 2018.

[4] Gerald Carlino and William R. Kerr. Agglomeration and Innovation. In *Handbook of Regional and Urban Economics*, volume 5, pages 349–404. 2015.

[5] Henry Chen, Paul Gompers, Anna Kovner, and Josh Lerner. Buy local? The geography of venture capital. *Journal of Urban Economics*, 67(1):90–102, 2010.

[6] Iain M. Cockburn, Rebecca Henderson, and Scott Stern. The Impact of Artificial Intelligence on Innovation. may 2018.

[7] Péter Érdi, Kinga Makovi, Zoltán Somogyvári, Katherine Strandburg, Jan Tobochnik, Péter Volf, and László Zalányi. Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics*, 95(1):225–242, 2013.

[8] Roberto Fontana, Alessandro Nuvolari, Hiroshi Shimizu, and Andrea Vezzulli. Reassessing patent propensity: Evidence from a dataset of R&D awards, 1977-2004. *Research Policy*, 42(10):1780–1792, 2013.

[9] Chris Forman, Avi Goldfarb, and Shane Greenstein. Agglomeration of Invention in the Bay Area: Not Just ICT. *American Economic Review*, 106(5):146–151, may 2016.

[10] Bronwyn Hall, Adam Jaffe, and Manuel Trajtenberg. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. oct 2001.

[11] Bronwyn H. Hall and Rosemarie Ham Ziedonis. The Patent Paradox Revisited : An Empirical Study of Patenting in the U . S . Semiconductor Industry , 1979-1995. 32(1):101–128, 2001.

[12] Jerry Hausman, Bronwyn H . Hall, and Zvi Griliches. Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica*, 52(4):909–938, 1984.

[13] Steven N Kaplan and Josh Lerner. Venture Capital Data: Opportunities and Challenges. 2016.

[14] Samuel Kortum and Josh Lerner. Does venture capital spur innovation? *RAND Journal of Economics*, 31:674–692, 2000.

[15] Saul Lach. Do R & D Subsidies Stimulate or Displace Private R & D ? Evidence from Israel. *The Journal of Industrial Economics*, 50(4):369–390, 2002.

[16] Alessandro Muscio, Davide Quaglione, and Giovanna Vallanti. Does government funding complement or substitute private research funding to universities? *Research Policy*, 42(1):63–75, 2013.

[17] Francis Narin, Kimberly S. Hamilton, and Dominic Olivastro. The increasing linkage between U.S. technology and public science. *Research Policy*, 12:317–330, 1997.

[18] Walter W. Powell, Kenneth W. Koput, James I. Bowie, and Laurel Smith-Doerr. The spatial clustering of science and capital: Accounting for biotech firm - Venture capital relationships. *Regional Studies*, 36(3):291–305, 2002.

[19] Marco Da Rin, Thomas F. Hellmann, and Manju Puri. A survey of venture capital research. 2011.

[20] Sampsa Samila and Olav Sorenson. Venture capital as a catalyst to commercialization. *Research Policy*, 39(10):1348–1360, dec 2010.

[21] Sampsa Samila and Olav Sorenson. Venture Capital, Entrepreneurship, and Economic Growth. *The Review of Economics and Statistics*, 93(1):338–349, 2011.

[22] F.M. Schefer. The prooensity to patent. *International Journal of Industrial Organization*, 1:107–128, 1983.

[23] Olav Sorenson and Toby Stuart. Syndication Networks and the Spatial Distribution of Venture Capital Investments. *American Journal of Sociology*, 106(6):1546–1588, 2001.

[24] Scott J. Wallsten. The Effects of Government-Industry R & D Programs on Private R & D : The Case of the Small Business Innovation Research Program. *The RAND Journal of Economics*,, 31(1):82–100, 2000.